



## **CRISPR-Cas12a**

### **functional overview and applications**

Paul, Bijoya; Montoya, Guillermo

*Published in:*  
Biomedical Journal

*DOI:*  
[10.1016/j.bj.2019.10.005](https://doi.org/10.1016/j.bj.2019.10.005)

*Publication date:*  
2020

*Document version*  
Publisher's PDF, also known as Version of record

*Document license:*  
[CC BY-NC-ND](#)

*Citation for published version (APA):*  
Paul, B., & Montoya, G. (2020). CRISPR-Cas12a: functional overview and applications. *Biomedical Journal*, 43(1), 8-17. <https://doi.org/10.1016/j.bj.2019.10.005>

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

ScienceDirect

Biomedical Journal

journal homepage: [www.elsevier.com/locate/bj](http://www.elsevier.com/locate/bj)

## Review Article

## CRISPR-Cas12a: Functional overview and applications

Bijoya Paul, Guillermo Montoya<sup>\*</sup>

Structural Molecular Biology Group, Novo Nordisk Foundation Centre for Protein Research, Faculty of Health and Medical Sciences University of Copenhagen, Copenhagen, Denmark



## ARTICLE INFO

## Article history:

Received 30 August 2019

Accepted 21 October 2019

Available online 5 February 2020

## Keywords:

CRISPR-Cas12a

RNA guided endonucleases

crRNA biogenesis

Indiscriminate ssDNAse

Endonuclease recycling

Genome editing

## ABSTRACT

Prokaryotes have developed an adaptive immune system called Clustered regularly interspaced short palindromic repeats (CRISPR) to combat attacks by foreign mobile genetic elements (MGEs) such as plasmids and phages. In the past decade, the widely characterized CRISPR-Cas9 enzyme has been redesigned to trigger a genome editing revolution. Class II type V CRISPR-Cas12a is a new RNA guided endonuclease that has been recently harnessed as an alternative genome editing tool, which is emerging as a powerful molecular scissor to consider in the genome editing application landscape. In this review, we aim to provide a mechanistic insight into the working mechanism of Cas12a, comparing it with Cas9, and eventually provide an overview of its current applications in genome editing and biotechnology applications.

Genome editing is a type of genetic engineering where a DNA is inserted, deleted or replaced in the genome of a living organism. The application of this technology has revolutionized various research areas ranging from biomedicine to biotechnology or synthetic biology [1–3]. A key point to initiate the editing is the need to generate a double strand break (DSB) in the DNA at a specific locus in the genome. To achieve this precise DSB researchers have developed engineered

nucleases, also termed “molecular scissors”. Previous efforts have focused in the molecular understanding and redesign of different protein templates, such as homing endonucleases, zinc finger nucleases (ZFN) and TALEN. These tools have shown their utility in different genome editing applications [4–6] including the correction of mutations involving monogenic diseases [7,8]. However, the engineering of new DNA specificities in these protein scaffolds is cumbersome.

<sup>\*</sup> Corresponding author. Structural Molecular Biology Group, Novo Nordisk Foundation Centre for Protein Research, Faculty of Health and Medical Sciences University of Copenhagen, Blegdamsvej 3-B, Copenhagen 2200, Denmark.

E-mail address: [guillermo.montoya@cpr.ku.dk](mailto:guillermo.montoya@cpr.ku.dk) (G. Montoya).

Peer review under responsibility of Chang Gung University.

<https://doi.org/10.1016/j.bj.2019.10.005>

2319-4170/© 2019 Chang Gung University. Publishing services by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Therefore, the development of the versatile CRISPR-Cas systems (Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR associated proteins) as molecular scissors, where a simple exchange of the RNA guide sequence is enough to redesign the nuclease specificity has paved the way for a revolution in the life sciences [9–11].

CRISPR repeats are associated with Cas proteins constituting an adaptive immune system in bacteria and archaea protecting them from foreign mobile genetic elements [12]. The discovery of CRISPR-Cas revealed the capability of bacteria and archaea to acquire and integrate genetic elements into its own genome, demonstrating the exchange of information between the environment and prokaryotic genomes [12]. The genetic record of previous attacks by foreign nucleic acids is stored in the CRISPR arrays. These arrays are made of short and conserved repetitive sequences called repeats which are strategically placed between unique sequences called spacers. They are inserted by specialized Cas proteins into the CRISPR array during infections by invading nucleic acids [13–16]. The adaptive immunity by prokaryotes against foreign MGEs is achieved through the formation of RNA-guided endonucleases, which constitute the effector complexes and are able to detect secondary infection by a foreign DNA that was previously incorporated into the CRISPR array [10].

The CRISPR-Cas systems are classified into two classes (Classes 1 and 2) that are subdivided into six types (types I through VI). Class 1 (types I, III and IV) systems use multiple Cas proteins in their CRISPR ribonucleoprotein effector nucleases and Class 2 systems (types II, V and VI) use a single Cas protein [17]. Class 1 CRISPR-Cas systems are most commonly found in bacteria and archaea, and comprise ~90% of all identified CRISPR-Cas loci. The Class 2 CRISPR-Cas systems, comprising the remaining ~10%, exists almost exclusively in bacteria [18], and assemble a ribonucleoprotein complex, consisting of a CRISPR RNA (crRNA) and a Cas protein [10]. The crRNA contains information to target a specific DNA sequence [19]. These multidomain effector proteins achieve interference by complementarity between the crRNA and the target sequence after recognition of the PAM (Protospacer Adjacent Motif) sequence, which is adjacent to the target DNA [20]. These ribonucleoprotein complexes have been redesigned for precise genome editing by providing a crRNA with a redesigned guide sequence, which is complementary to the sequence of the targeted DNA [21,22]. The most widely characterised CRISPR-Cas system is the type II subtype II-A that is found in *Streptococcus pyogenes* (Sp), which uses the protein SpCas9. Cas9 was the first Cas-protein engineered for use in gene editing [9]. Class 2 type V is further classified into 4 subtypes (V-A, V-B, V-C, V-U). At present, V-C and V-U remain widely uncharacterised and no structural information on these systems is available [23]. V-A encodes the protein Cas12a (also known as Cpf1) and recently several high resolution structures of Cas12a have provided an insight into its working mechanism [24–26].

This system, involving RNA-guided interference, has been harnessed into a versatile biotechnological tool for genome editing [9–11], whereby a simple exchange of the RNA guide sequence can be employed to re-engineer nuclease specificity, leading to a revolution in the life sciences. Cas9, belonging to

Class 2 Type II CRISPR-Cas interference system, is the more extensively used tool for genome editing. However, the overwhelming efficiency in genome sequencing of different organisms has generated a large amount of data helping the identification of new systems whose use in genome editing is currently being explored. Among them there are new members of Class 2, such as Cas12a [18,23,27].

CRISPR-Cas immunity involves three major sequential steps: *adaptation*, *expression/maturation* and *interference* (Fig. 1), each step needs specific Cas proteins encoded by the *cas* genes near the CRISPR array, together with other accessory proteins [2,28,29]. The CRISPR-Cas adaptation stage involves the identification and extraction of the protospacer from the invading DNA/RNA and its subsequent incorporation into the CRISPR array. Both these functions are performed by the versatile Cas1-Cas2 adaptation complex. The identification of the protospacer starts with the recognition of the PAM by the adaptation complex [28], subsequently the spacer (sequence adjacent to the PAM) is integrated into the CRISPR array and the conserved repeat sequence is duplicated. The PAM sequence is excluded from the CRISPR array and is one of the first recognition motifs used for identifying target nucleic acids for degradation [16,24,28]. During the expression/maturation stage, the CRISPR array is transcribed into a long pre-CRISPR RNA (pre-crRNA) molecule. The pre-crRNA is processed into shorter crRNA molecules each containing a spacer and a part of the repeat sequence. Finally, interference can occur, after the crRNA forms a complex with the effector protein, forming a functional RNA guided endonuclease. This endonuclease is guided by the crRNA, which after PAM recognition hybridizes with the target DNA through its spacer sequence, and eventually, cuts the target DNA sequence. In this review, we describe the structural and functional features of Cas12a, a cousin of Cas9 belonging to the Class 2 Type V CRISPR-Cas system, which has been repurposed into an alternative and promising gene editing tool based on its substantial differences with Cas9 [17,30].

## Cas12a crRNA biogenesis

RNA sequencing of small RNA molecules extracted from *Francisella novicida* U112 culture containing Cas12a-based CRISPR loci revealed that mature crRNAs for Cas12a are 42–44 nt in length, with the first 19/20 nt corresponding to the repeat sequence and the remaining 23–25 nt to the spacer sequence [31]. In type II CRISPR systems, the maturation of crRNA is done by host housekeeping protein RNase III together with the trans-activating crRNA (tracrRNA), which is base paired with the pre-crRNA, in presence of Cas9 [32,33]. In contrast, it has been shown that Cas12a processes its own pre-crRNA into mature crRNAs, without the requirement of a tracrRNA, making it a unique effector protein with both endoribonuclease and endonuclease activities [34]. After the pre-crRNA has been transcribed during the expression stage, Cas12a cuts it 4 nt upstream of the hairpin structures formed by the CRISPR repeats, producing intermediate crRNA molecules which undergo further processing *in vivo* into mature crRNAs.

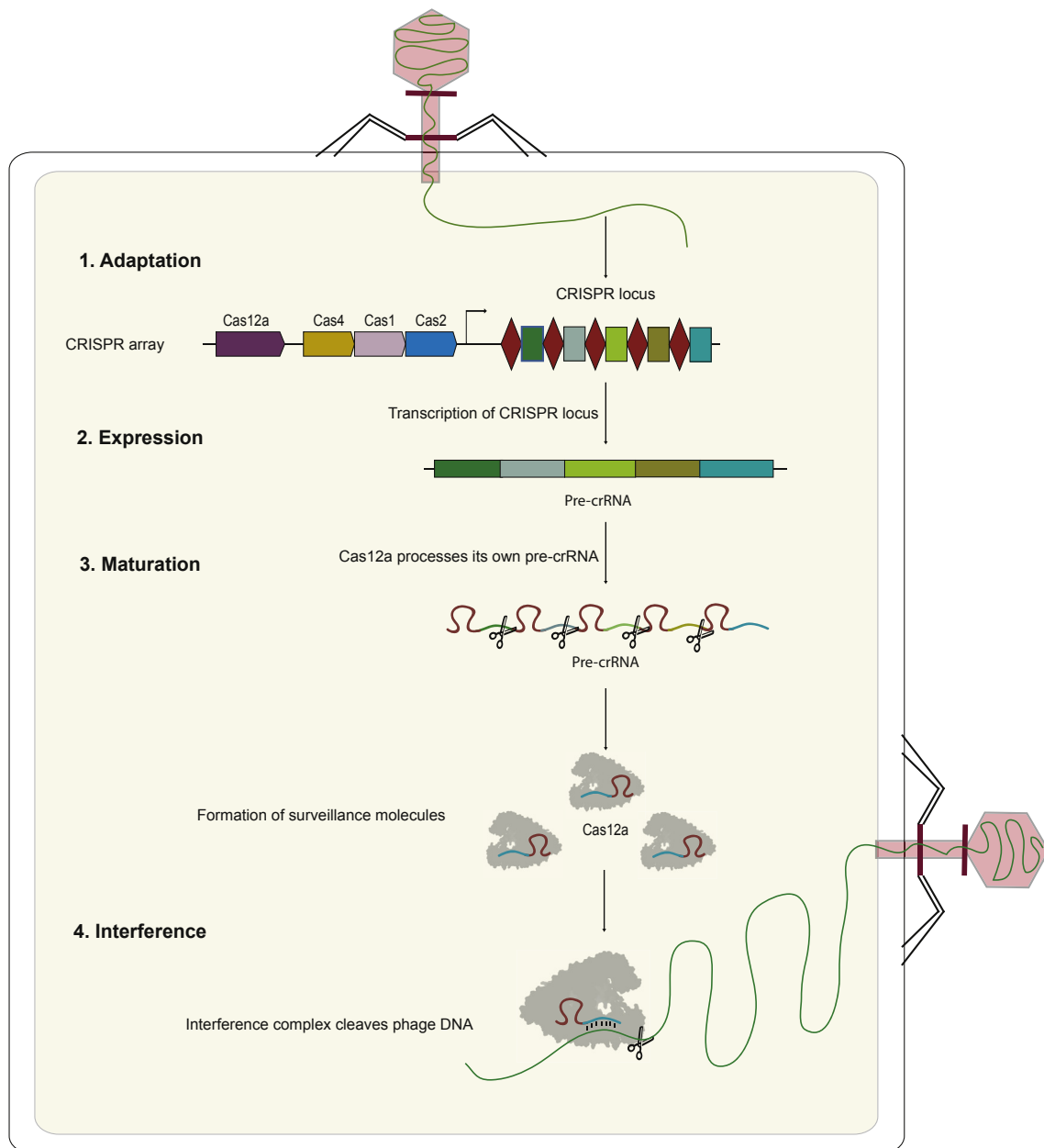


Fig. 1 Stages of CRISPR-Cas immunity- Adaptation, Expression/Maturation, Interference.

### Cas12a domain organisation and ribonucleoprotein complex assembly

Type II (Cas9) and type V (Cas12a) CRISPR-Cas systems possess a characteristic Ruv-C like nuclease domain (Fig. 2A), which has been shown to be related to IS605 family transposon encoded TnpB proteins [18]. Crystallographic and cryo-EM data [24,31,35,36] reveal that Cas12a adopts a bilobed structure formed by the REC and Nuc lobes (Fig. 2B). The REC lobe is comprised of REC1 and REC2 domains, and the Nuc lobe is comprised of the RuvC, the PAM-interacting (PI) and the WED domains, and additionally, the bridge helix (BH). The RuvC endonuclease domain of this effector

protein is made up of three discontinuous parts (RuvC I-III). The RNase site for processing its own crRNA is situated in the WED-III subdomain, and the DNase site is located in the interface between the RuvC and the Nuc domains. These structural studies have also shown that the only the 5' repeat region of the crRNA is involved in the assembly of the binary complex. The 19/20 nt repeat region forms a pseudoknot structure through intramolecular base pairing. The crRNA is stabilized through interactions with the WED, RuvC and REC2 domains of the endonuclease, as well as two hydrated  $Mg^{2+}$  ions [24–26,35]. This binary interference complex is then responsible for recognizing and degrading foreign DNA.

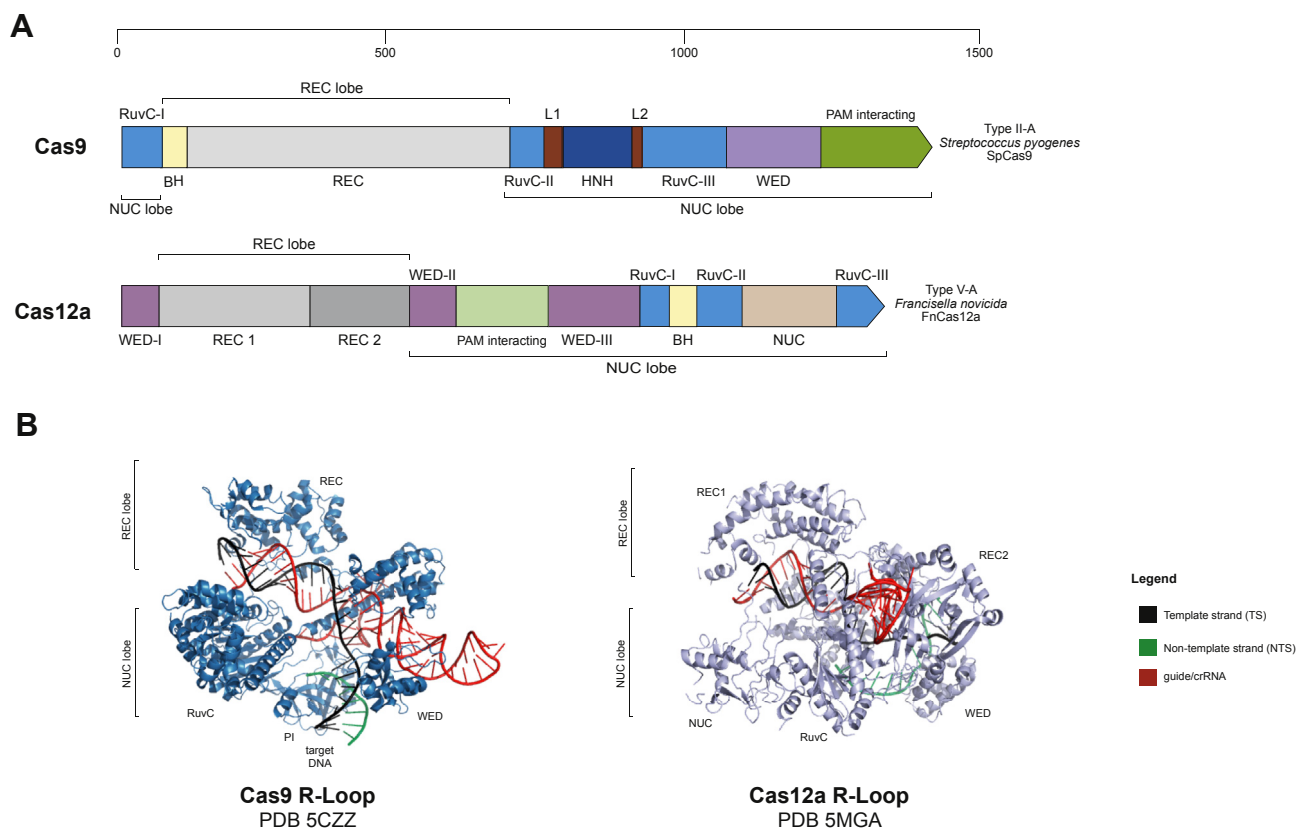


Fig. 2 Domain organisation and R-loop complex of Ca9 and Cas12a.

## PAM recognition

PAM recognition is a critical initial step in identifying a prospective DNA molecule for degradation since the PAM allows the CRISPR-Cas systems to distinguish their own genomic DNA from invading nucleic acids [37]. Cas12a employs a multistep quality control mechanism to ensure the accurate and precise recognition of target spacer sequences. The WED II-III, REC1 and PAM-interacting domains are responsible for PAM recognition and for initiating the hybridization of the DNA target with the crRNA. After recognition of the dsDNA by WED and REC1 domains, the conserved loop-lysine helix-loop (LKL) region in the PI domain, containing three conserved lysines (K667, K671, K677 in FnCas12a), inserts the helix into the PAM duplex with assistance from two conserved prolines in the LKL region. Structural studies show the helix is inserted at an angle of 45° with respect to the dsDNA longitudinal axis, promoting the unwinding of the helical dsDNA. The critical positioning of the three conserved lysines on the dsDNA initiates the uncoupling of the Watson–Crick interaction between the base pairs of the dsDNA after the PAM. The target dsDNA unzipping allows the hybridization of the crRNA with the strand containing the PAM, the ‘target strand (TS)’, while the uncoupled DNA strand, non-target strand (NTS), is conducted towards the DNase site by the PAM-interacting domain [24,26,35]. Cas12a has been shown to efficiently target spacer sequences following 5'-T-rich PAM sequence. The PAM for LbCas12a and AsCas12a has a sequence of 5'-TTTN-3' and for

FnCas12a a sequence of 5'-TTN-3' and is situated upstream of the 5' end of the non-target strand [26,31,34]. It has also been shown that in addition to the canonical 5'-TTTN-3' PAM, Cas12a also exhibits relaxed PAM recognition for suboptimal C-containing PAM sequences by forming altered interactions with the targeted DNA duplex [38].

## DNA unzipping, propagation and cleavage

Once the crRNA-DNA hybrid R-loop starts forming, the enzyme then looks for a seed sequence of 3–5 nt on the PAM proximal end, the next check point in the correct identification of the target (Fig. 3). It has been reported that mismatches in the seed sequence results in the loss of cleavage activity [31,34]. Presence of the seed sequences promotes further hybridization of the crRNA-target DNA. Structural studies have shown that TS and NTS follow different pathways to the nuclease site [24], with several residues in the PI domain undergoing conformational changes and adopting a ‘rail’ shape to accommodate the nt-strand and eventually guiding it to the catalytic site. This structure also shows the presence of a barrier, the septum, to prevent the re-annealing of the dsDNA.

A recent cryo-EM analysis on the intermediate catalysis products of Cas12a [36] revealed that the enzyme utilizes a further three-checkpoint control to sense the hybridization between the crRNA and the DNA. Three regions within the enzyme, the ‘REC linker’, the ‘lid’ and the ‘REC finger’

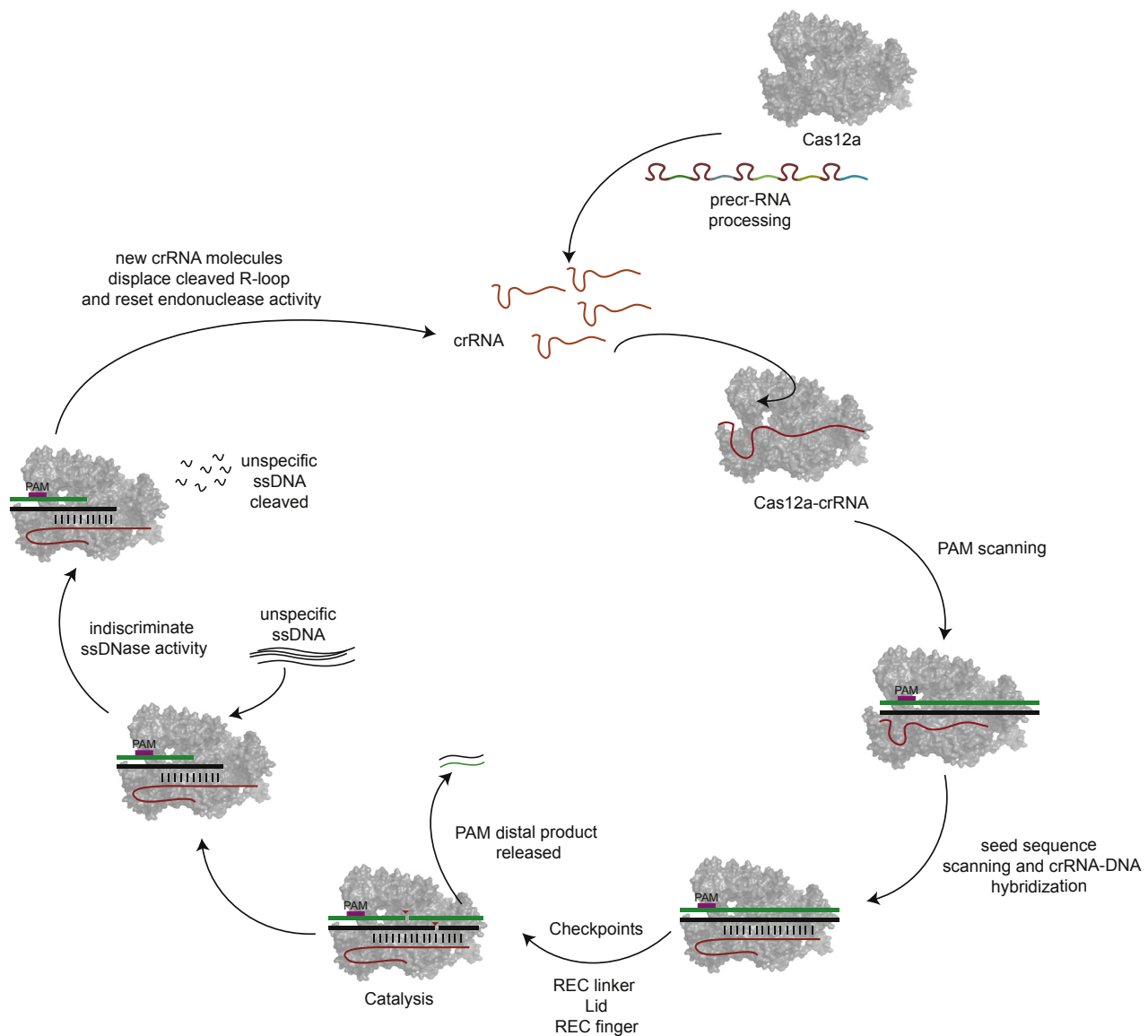


Fig. 3 Model mapping the catalytic pathway of CRISPR-Cas12a.

sequentially scan the hybrid through conformational changes and only when all three checkpoints are able to recognize the hybrid, the enzyme is in a conformation competent for catalysis. The endonuclease produces a staggered cut on a PAM distal site on the DNA with a 5 nt overhang on the target strand [31,36], and the PAM distal end of the cleaved product is then released from the complex [39].

Cleavage in the t-strand of the DNA by Cas12a produces a 5'-phosphorylated product [35]. In order for both DNA strands to be cut, they must enter the catalytic site with a 5'-3' polarity. Structural studies reveal that the NTS is positioned to enter the RuvC-Nuc pocket with the 5'-3' polarity, while the TS has the reverse polarity. An smFRET analysis suggests that Cas12a has to undergo conformational changes in the distal part of the REC and NUC lobes in order to allow the TS enter the nuclease site with the correct polarity [36]. This could explain why the NTS appears to be hydrolyzed faster than the TS. Therefore, the cleavage of the NTS is a consequence of the

proper positioning of this strand in the RuvC-Nuc catalytic pocket rather than a requirement to initiate the cleavage reaction. After both strands have been cleaved, the PAM distal end of the cleavage product dissociates from the complex, but the PAM proximal site remains associated to Cas12a forming a cleaved R-loop [24,39].

### Indiscriminate ssDNA cleavage

Besides high-specific dsDNA cleavage, Cas12a has also been shown to exhibit indiscriminate ssDNA degradation activity upon activation with a ssDNA complementary to the crRNA guide. This activity is displayed by all Cas12a orthologs and degrades any available ssDNA molecule into single/double nucleotides [40]. Comparisons of the structures of Cas12a before, during and after cleavage reveal the structural changes that result in such an indiscriminate activity. The



lid region, which is involved in the checkpoints for accurate target recognition [36] is responsible for this action. Before the crRNA-DNA hybrid is formed, the lid occludes the cleft where the catalytic residues reside. Upon formation of the hybrid, the lid changes conformation to form an  $\alpha$  helix, thus interacting with the crRNA of the hybrid assembly, thus dissociating the polar interactions and making available the catalytic pocket. In the R-loop structure after cleavage [24], this region appears disordered indicating that the catalytic site is accessible after the distal part of the dsDNA substrate dissociates from the complex. Therefore, the catalytic cleft is open and able to sever ssDNA indiscriminately. This molecular mechanism would explain how ssDNA molecules are degraded by Cas12a after being activated by the presence of the RNA-DNA hybrid [24,35,36]. In addition, recent studies have reported non-specific nicking of target sequences bearing mismatches in distal regions of the target DNA [41], suggesting that this could be a problem for potential applications.

### Cas12a endonuclease recycling

Unspecific ssDNA degradation presents a potential harmful situation for the host cell since it could hinder basic cellular processes such as replication, transcription and DNA repair. This poses an important question: how could we eliminate this harmful indiscriminate activity? As the cell cannot allow an indiscriminate ssDNA degradation unleashed. The answer to these problems can be found in the bacterial genomes encoding Cas12a. Use of a conserved sequence of the crRNA for a database search disclosed that different bacteria encode a single copy of the Cas12a gene, whereas they encode multiple copies (up to 68) of the crRNA [18]. If the transcription rates are presumed to be similar, at any given time, the concentration of various crRNAs in the cell would be multiple times higher than that of the enzyme. It has been experimentally shown that with sufficient concentration of a new crRNA molecule, it is able to displace the cleaved R-loop from the enzyme, with the help of accessory host proteins, forming a new interference complex (Fig. 3). This shows that Cas12a can revert the active conformation to shut down unspecific activity by displacing the cleaved R-loop with a new crRNA. In doing so, it reverts back to a conformation where the molecular 'lid' forms polar interactions again to make the catalytic pocket inaccessible [36]. By doing so, not only does the endonuclease shut down the indiscriminate ssDNase activity, but also recycles its catalytic activity towards other target DNAs.

### Cas12a vs Cas9

Cas12a and Cas9 have striking functional similarities despite having evolved through independent pathways (Figs. 2 and 4), with similar sizes (1368 amino acids for SpCas9; 1307 for FnCas12a). They are both multidomain effector proteins and adopt a bilobed architecture when in complex with their respective RNAs. Cas9 requires two RNA molecules: tracrRNA and a crRNA, whereas Cas12a requires only a single RNA molecule, the crRNA. Cas9 possesses two nuclease sites HNH

and RuvC domains, while Cas12a possesses only one nuclease site in the RuvC domain. Additionally, Cas12a also possesses an RNA processing site [1,9,24,31,34,37]. There are distinct differences in the mechanisms employed by the two proteins when it comes to RNA processing, PAM recognition, target DNA binding and eventually catalysis.

After the CRISPR array has been transcribed into a long pre-crRNA molecule, it is processed into mature crRNAs before it can form an RNP with the endonuclease. In the case of Cas9, the tracrRNA (encoded close to the CRISPR locus) first needs to hybridize with the pre-crRNA, and then this hybrid RNA-RNA duplex structure is recognized by Cas9, following which the host RNase III cleaves the duplex, leaving a ~75 nt long tracrRNA and a 39–42 nt long crRNA, which then forms the RNP complex responsible for recognizing and degrading the target DNA [33,42,43]. In contrast, Cas12a does not require a tracrRNA or RNase III, since the protein processes its own crRNA in its ribonuclease catalytic site [44]. In type-V CRISPR locus of *F. novicida*, the spacers are 27–32bp sequences interspersed by 36bp repeat sequences in between the spacers. In the entire CRISPR array transcript, which is the pre-crRNA, the repeat derived sequences form pseudoknots, which are recognized by Cas12a. Following the recognition, the pre-crRNA is cleaved forming ~43 nt mature crRNA [25,26,31,34,35,45]. Logistically, Cas12a presents a more minimalistic system than Cas9.

For Cas9 targeted DNA sequences, the PAM is situated downstream of the spacer sequence on the non-template strand, and is recognized by the PI domain, which is primed for identifying a 5'-NGG-3' PAM [43,46]. In contrast, Cas12a recognizes A-T rich sequences, with the PAM, typically 5'-TTTV-3', located upstream of the spacer. Upon PAM recognition, the target DNA is unzipped and hybridization of the RNA-DNA takes place. For both enzymes exists a crucial seed sequence next to the PAM to determine the specificity of target DNA binding. The seed sequence for Cas9 is about ~10 nt whereas for Cas12a it is about ~5–6 nt [9,31,34,35]. When the hybridization of the DNA with the RNA is complete, Cas9 cleaves the template strand and the non-template strands in the catalytic sites located in the HNH and the RuvC domains respectively, producing a blunt DSB, with the cleavage site being 3 base pairs upstream from the PAM sequence [9]. However, in the case of Cas12a due to the presence of a single nuclease site, the strands of the DNA are cut in the same nuclease site. Since it has been shown that the two strands follow different pathways to reach the catalytic site, explaining the staggered DSB produced by Cas12a [24,31].

Cleavage fidelity is an important issue for many of the nucleases used in genome editing applications [47]. An optimal tool must introduce modifications just on the target site, leaving the rest of the genome unmodified in order to avoid undesired changes in other sites of the genome with unpredictable consequences. Therefore, specificity and the resulting cleavage products are key in genome modification applications. A recent comparison of different Cas12a and Cas9 from different species using nuclease digestion and deep sequencing (NucleaSeq) *in vitro*, revealed that both enzymes share similar types of specificities and tolerate similar mismatches [48], in contrast to *in vivo* reports that show the lower off-target effects of Cas12a [49,50]. This apparent



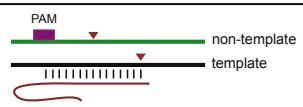
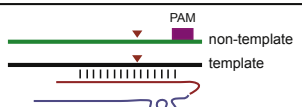
	Cas12a	Cas9
Size of protein	~1300 amino acids	~1000-1600 amino acids
RNA	 crRNA Single RNA molecule	 crRNA tracrRNA Two RNA molecules
Nuclease sites	Single nuclease site RuvC-Nuc	2 nuclease domains HNH and RuvC
Type of cut	 non-template template Staggered ends	 non-template template Blunt ends
PAM requirements	Recognises 5' T-rich PAM sequences of 3-4 nt	Recognises 3' G-rich PAM sequences of 3-5 nt
precrRNA processing	possesses intrinsic RNase activity to process precr-RNA	requires host RNase III and tracrRNA

Fig. 4 Cas12a vs Cas9.

contradiction may be related to different recognition and cleavage kinetics, but also to a possible different behaviour of these cutters on a chromatin context, thus posing the question whether *in vitro* or *in vivo* approaches should be pursued for nuclease redesign efforts.

### CRISPR-Cas12a mediated genome editing

Application of CRISPR-Cas systems as molecular tools for genome editing exploits their ability to produce a double strand break (DSB) at a specific genomic locus, and depends entirely on the host cell DNA repair machinery to fix the lesion produced by these systems. The repair mechanisms can be either of the following processes: homology-directed repair (HDR) or non-homologous end joining (NHEJ). HDR utilizes a template DNA that is homologous to the break site (an unbroken sister chromatid or a homologous chromosome) to repair the DSB, whereas NHEJ is based on direct joining of broken ends of the DSB, making NHEJ the more error prone mechanism of the two. HDR can thus be used to supply exogenous template DNA to implement a user defined change in the host genome. NHEJ can be applied for gene disruption

whereas HDR allows for the scope of introducing new genetic information or direct correction of the sequence at a specific locus.

At the center of CRISPR mediated genome engineering today is Cas9, with applications including, but not limited to, gene knockout and precise genome editing. Despite the rapid advances in genome editing by Cas9, it still presents challenges owing to the possibility of off-target effects and difficulty of delivering the ribonucleoprotein particle [18]. Cas12a, owing to its substantial differences with Cas9, presents an alternate molecular genome editing tool. The use of Cas12a in genome editing for various cell types has been probed in several studies up to date. Comparative studies of gene repression by catalytically dead Cas9 from *S. pyogenes* (SpdCas9) and catalytically dead Cas12a from *Eubacterium eligens* (EedCas12a) revealed that the latter displays a higher gene repression in the template strand of the target DNA than SpdCas9 [51]. It was also shown that the pre-crRNA processing activity of Cas12a makes it an attractive candidate for multiplex gene regulation, which is cumbersome when attempted with Cas9 [52]. This auto-processing of its own crRNA has been used to modify multiple genetic elements simultaneously generating constitutive, conditional, inducible, orthogonal and multiplexed genome



engineering of endogenous targets using multiple CRISPR RNAs delivered on a single plasmid [53].

The viability of this approach has been further established by other studies, in which multiplex gene regulation by Cas12a was successfully observed in bacteria, plants, as well as in mammalian cells [52,54–56]. Cas12a can also serve as a solution in cell types where use of Cas9 is toxic, such as in some industrial strains of *Streptomyces* [54].

Targeted mutagenesis in plants can also be achieved through co-expression of Cas12a and its cognate crRNA *in vivo*, as was shown in rice. Additionally, it was also shown that the mutagenesis was more efficient through the use of pre-crRNAs with full-length direct repeat sequences than with mature crRNAs [57]. Efficient mutagenesis through delivery of the pre-assembled ribonucleoprotein (RNP) particle was also observed in soybean and wild tobacco. The RNP was assembled from recombinantly expressed Cas12a and *in vitro* transcribed or chemically synthesized crRNAs [58].

Successful gene editing of mammalian cells using Cas12a include correction of mutations causing Duchenne muscular dystrophy (DMD) in patient derived induced pluripotent stem cells (iPSCs) and in *mdx* mice, a popular model for studying DMD. Dystrophin expression was reinstated in iPSCs after Cas12a-mediated gene editing, while in the *mdx* mice, corrections in the pathophysiological hallmarks of muscular dystrophy were observed [59]. Delivery of the adenovirus vector with an AsCas12a expression cassette yielded successful mutations in primary human hepatocytes from humanized mice with chimeric liver [60]. Cas12a-mediated genome editing was also used to engineer rat models that mimic human atherosclerosis and this system may have potential applications in understanding early stage atherosclerosis [61].

All of the above studies show Cas12a can be engineered for various applications. Despite the numerous recent advances in the application of Cas12a, there remain vast avenues of unexplored potential of Cas12a in terms of therapeutics and diagnostics.

## Cas12a applications in bioengineering

Currently, a vast effort is ongoing to redesign all these tools for biomedical and biotechnological applications. However, recent studies have envisioned the possibility of using CRISPR-Cas nucleases in bioengineering of smart materials, for example hydrogels [62]. These water-filled polymers are encapsulated by DNA. In a recent study, Cas12a has been used to specifically degrade the DNA scaffold of DNA hydrogels, thus opening the possibility that this smart cutter can be turned into a programmable device to deliver the cargo of DNA encaged hydrogels in a determined location at a certain time. The cleavage properties of Cas12a make it an ideal candidate to promote controlled delivery of the cargo. Although, application of these approaches and their combinations can be now envisioned by many researchers, the range of possibilities in different areas is so large that it is beyond our imagination.

## Conclusions

In this review we have sought to offer a condensed overview of the functionality of CRISPR-Cas12a, discussing the structural and functional features of the different stages of the reaction pathway leading up to the catalysis of the target DNA, and eventually discussing the applications of Cas12a in brief. Cas12a employs a multi-checkpoint mechanism to ensure precise targeting of DNA, which is a desirable property in a genome editing tool in order to have low off target effects. Although it has been shown that the indiscriminate ssDNA degradation of Cas12a could be shut down through the recruitment of a new crRNA molecule, it could still potentially harm the host cell targeted for genome modification. Modulation of this activity is necessary to achieve higher regulation and control of Cas12a catalysis, and in turn to achieve a more robust genome editing tool. In this direction structural information has been used to redesign Cas12a obtaining variants without ssDNA unspecific activity, thus severing only dsDNA specifically [36].

Currently, Cas9 and Cas12a, are the sole members of the CRISPR family that have been utilized for genome editing. Owing to their significant similarities and differences, just these two endonucleases between themselves have made the applications of CRISPR highly versatile. Cas12a, in some cases, offers certain advantages over Cas9, for example in its capability to be used for multiplex genome editing and production of staggered DSB, which promotes HDR instead of NHEJ. Significant research also is ongoing to engineer artificial variants of Cas9 and Cas12a to recognize different PAM than the wild type proteins, which will facilitate the targeting of a wider library of genomes.

The rapid advent of the CRISPR-Cas technology for genome manipulation has been revolutionary for life sciences. Despite the vast application areas of this technology, the current state of the art of the CRISPR molecular tools (Cas9 or Cas12a) suffers from one important drawback: dependence on host cell DNA repair machinery. Both Cas9 and Cas12a based technology produce a double strand break (DSB) in the target DNA, and this break is then repaired by endogenous DNA repair machinery with or without the presence of a template. Although these tools have been successfully utilized to obtain precise insertion of DNA into the targeted genomic loci, their efficiency differs from cell type to cell type [63–65]. DNA repair through HDR is also related to active cell division, which makes these tools ineffective in cell types that are not actively dividing, such as neurons. Recent studies characterizing CRISPR-associated transposase (CAST), which comprises Tn7-like transposase subunits and a CRISPR effector from type V–K, could pave the way to new avenues of gene editing using CRISPR systems since these systems are self-sufficient in precise DNA insertion and do not depend on endogenous cell DNA repair machinery [66,67]. However, a large ongoing research is aiming to tailor both Cas9 and Cas12a further to ensure precise DNA insertion into the targeted genome. Even apart from this apparent drawback, both these tools have a vast range of applicability and ongoing efforts are

striving to produce improved and more robust engineered genome editing tools.

## Conflicts of interest

GM declares that he is an inventor in a Cas12a patent application.

## Acknowledgements

The Novo Nordisk Foundation Center for Protein Research is supported financially by the Novo Nordisk Foundation (grant NNF14CC0001). Our work is also supported by the cryoEM (grant NNF0024386), cryoNET (grant NNF17SA0030214), and Distinguished Investigator (NNF18OC0055061) grants to GM, who is a member of the Integrative Structural Biology Cluster (ISBUC) at the University of Copenhagen. We thank Stefano Stella for advice during the preparation of the manuscript.

## REFERENCES

- [1] Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. *Science* 2014;346:1258096.
- [2] Wright AV, Nuñez JK, Doudna JA. Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* 2016;164:29–44.
- [3] Barrangou R, Horvath P. A decade of discovery: CRISPR functions and applications. *Nat Microbiol* 2017;2:17092.
- [4] Leitao AL, Costa MC, Enguita FJ. Applications of genome editing by programmable nucleases to the metabolic engineering of secondary metabolites. *J Biotechnol* 2017;241:50–60.
- [5] Knipping F, Osborn MJ, Petri K, Tolar J, Glimm H, von Kalle C, et al. Genome-wide specificity of highly efficient TALENs and CRISPR/Cas9 for T cell receptor modification. *Mol Ther - Methods Clin Dev* 2017;4:213–24.
- [6] Chandrasegaran S, Carroll D. Origins of programmable nucleases for genome engineering. *J Mol Biol* 2016;428:963–89.
- [7] Tebas P, Stein D, Tang WW, Frank I, Wang SQ, Lee G, et al. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N Engl J Med* 2014;370:901–10.
- [8] Redondo P, Prieto J, Munoz IG, Alibes A, Stricher F, Serrano L, et al. Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases. *Nature* 2008;456:107–11.
- [9] Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012;337:816–21.
- [10] Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci* 2012;109:E2579–86.
- [11] Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/cas systems. *Science* 2013;339:819–23.
- [12] Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 2005;60:174–82.
- [13] Nuñez JK, Kranzusch PJ, Noeske J, Wright AV, Davies CW, Doudna JA. Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nat Struct Mol Biol* 2014;21:528–34.
- [14] Nuñez JK, Harrington LB, Kranzusch PJ, Engelman AN, Doudna JA. Foreign DNA capture during CRISPR–Cas adaptive immunity. *Nature* 2015;527:535–8.
- [15] Jackson SA, McKenzie RE, Fagerlund RD, Kieper SN, Fineran PC, Brouns SJJ. CRISPR-Cas: adapting to change. *Science* 2017;356:eaal5056.
- [16] Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 2009;155:733–40.
- [17] Nishimasu H, Nureki O. Structures and mechanisms of CRISPR RNA-guided effector nucleases. *Curr Opin Struct Biol* 2017;43:68–78.
- [18] Shmakov S, Smargon A, Scott D, Cox D, Pyzocha N, Yan W, et al. Diversity and evolution of class 2 CRISPR–Cas systems. *Nat Rev Microbiol* 2017;15:169–82.
- [19] Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* 2010;327:167–70.
- [20] Stella S, Alcón, Montoya G. Class2 CRISPR-cas RNA-guided endonucleases: Swiss army knives of genome editing. *Nat Struct Mol Biol* 2017;24:882–92.
- [21] Liu L, Chen P, Wang M, Li X, Wang J, Yin M, et al. C2c1-sgRNA complex structure reveals RNA-guided DNA cleavage mechanism. *Mol Cell* 2017;65:310–22.
- [22] Yang H, Gao P, Rajashankar KR, Patel DJ. PAM-dependent target DNA recognition and cleavage by C2c1 CRISPR-cas endonuclease. *Cell* 2017;167:1814–28. e12.
- [23] Shmakov S, Abudayyeh OOO, Makarova KSS, Wolf YII, Gootenberg JSS, Semenova E, et al. Discovery and functional characterization of diverse class 2 CRISPR-cas systems. *Mol Cell* 2015;60:385–97.
- [24] Stella S, Alcón P, Montoya G. Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature* 2017;546:559–63.
- [25] Dong D, Ren K, Qiu X, Zheng J, Guo M, Guan X, et al. The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* 2016;532:522–6.
- [26] Yamano T, Nishimasu H, Zetsche B, Hirano H, Slaymaker IM, Li Y, et al. Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* 2016;165:949–62.
- [27] Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, et al. An updated evolutionary classification of CRISPR–Cas systems. *Nat Rev Microbiol* 2015;13:722–36.
- [28] Amitai G, Sorek R. CRISPR–Cas adaptation: insights into the mechanism of action. *Nat Rev Microbiol* 2016;14:67–76.
- [29] van der Oost J, Westra ER, Jackson RN, Wiedenheft B. Unravelling the structural and mechanistic basis of CRISPR–Cas systems. *Nat Rev Microbiol* 2014;12:479–92.
- [30] Stella S, Alcón P, Montoya G. Class 2 CRISPR–Cas RNA-guided endonucleases: Swiss Army knives of genome editing. *Nat Struct Mol Biol* 2017;24:882–92.
- [31] Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 2015;163:759–71.
- [32] Charpentier E, Richter H, van der Oost J, White MF. Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiol Rev* 2015;39:428–41.
- [33] Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 2011;471:602–7.

- [34] Fonfara I, Richter H, Bratović M, Le Rhun A, Charpentier E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* 2016;532:517–21.
- [35] Swarts DC, van der Oost J, Jinek M. Structural basis for guide RNA processing and seed-dependent DNA targeting by CRISPR-cas12a. *Mol Cell* 2017;66:221–33. e4.
- [36] Stella S, Mesa P, Thomsen J, Paul B, Alcón P, Jensen SB, et al. Conformational activation promotes CRISPR-cas12a catalysis and resetting of the endonuclease activity. *Cell* 2018;175:1856–71. e21.
- [37] Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 2014;507:62–7.
- [38] Yamano T, Zetsche B, Ishitani R, Zhang F, Nishimasu H, Nureki O. Structural basis for the canonical and non-canonical PAM recognition by CRISPR-cpf1. *Mol Cell* 2017;67:633–45. e3.
- [39] Singh D, Mallon J, Poddar A, Wang Y, Tippana R, Yang O, et al. Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 (Cas12a). *Proc Natl Acad Sci* 2018;115:5444–9.
- [40] Chen JS, Ma E, Harrington LB, Da Costa M, Tian X, Palefsky JM, et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* 2018;360:436–9.
- [41] Murugan K, Seetharam AS, Severin AJ, Sashital DG. Pervasive off-target and double-stranded DNA nicking by CRISPR-Cas12a. *BioRxiv* 2019:657791.
- [42] Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ, et al. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* 2016;351:867–71.
- [43] Jiang F, Zhou K, Ma L, Gressel S, Doudna JA. A Cas9–guide RNA complex preorganized for target DNA recognition. *Science* 2015;348:1477–81.
- [44] Fonfara I, Richter H, Bratović M, Le Rhun A, Charpentier E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* 2016;532:517–21.
- [45] Gao P, Yang H, Rajashankar KR, Huang Z, Patel DJ. Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell Res* 2016;26:901–13.
- [46] Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 2014;513:569–73.
- [47] Stella S, Montoya G. The genome editing revolution: a CRISPR-Cas TALE off-target story. *Bioessays* 2016;38:S4–13.
- [48] Jones SK, Hawkins JA, Johnson NV, Jung C, Hu K, Rybarski JR, et al. Massively parallel kinetic profiling of natural and engineered CRISPR. nucleases. *BioRxiv* 2019:696393.
- [49] Kim D, Kim J, Hur JK, Been KW, Yoon S, Kim JS. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat Biotechnol* 2016;34:863–8.
- [50] Gao L, Cox DBT, Yan WX, Manteiga JC, Schneider MW, Yamano T, et al. Engineered Cpf1 variants with altered PAM specificities. *Nat Biotechnol* 2017;35:789–92.
- [51] Kim SK, Kim H, Ahn WC, Park KH, Woo EJ, Lee DH, et al. Efficient transcriptional gene repression by type V-A CRISPR-cpf1 from *Eubacterium eligens*. *ACS Synth Biol* 2017;6:1273–82.
- [52] Zetsche B, Heidenreich M, Mohanraju P, Fedorova I, Kneppers J, DeGennaro EM, et al. Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat Biotechnol* 2017;35:31–4.
- [53] Campa CC, Weisbach NR, Santinha AJ, Incarnato D, Platt RJ. Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. *Nat Methods* 2019;67:887–93.
- [54] Li L, Wei K, Zheng G, Liu X, Chen S, Jiang W, et al. CRISPR-Cpf1-Assisted multiplex genome editing and transcriptional repression in *Streptomyces*. *Appl Environ Microbiol* 2018;84:e00827–918.
- [55] Wang M, Mao Y, Lu Y, Tao X, Zhu J kang. Multiplex gene editing in rice using the CRISPR-cpf1 system. *Mol Plant* 2017;10:1011–3.
- [56] Zhang X, Wang J, Cheng Q, Zheng X, Zhao G, Wang J. Multiplexed gene regulation by CRISPR-ddCpf1. *Cell Discov* 2017;3:17018.
- [57] Xu R, Qin R, Li H, Li D, Li L, Wei P, et al. Generation of targeted mutant rice using a CRISPR-Cpf1 system. *Plant Biotechnol J* 2017;15:713–7.
- [58] Kim H, Kim ST, Ryu J, Kang BC, Kim JS, Kim SG. CRISPR/Cpf1-mediated DNA-free plant genome editing. *Nat Commun* 2017;8:14406.
- [59] Zhang Y, Long C, Li H, McAnally JR, Baskin KK, Shelton JM, et al. CRISPR-Cpf1 correction of muscular dystrophy mutations in human cardiomyocytes and mice. *Sci Adv* 2017;3:e1602814.
- [60] Tsukamoto T, Sakai E, Iizuka S, Taracena-Gándara M, Sakurai F, Mizuguchi H. Generation of the adenovirus vector-mediated CRISPR/Cpf1 system and the application for primary human hepatocytes prepared from humanized mice with chimeric liver. *Biol Pharm Bull* 2018;41:1089–95.
- [61] Lee JG, Ha CH, Yoon B, Cheong S-A, Kim G, Lee DJ, et al. Knockout rat models mimicking human atherosclerosis created by Cpf1-mediated gene targeting. *Sci Rep* 2019;9:2628.
- [62] English MA, Soenksen LR, Gayet RV, de Puig H, Angenent-Mari NM, Mao AS, et al. Programmable CRISPR-responsive smart materials. *Science* 2019;365:780–5.
- [63] Jasin M, Rothstein R. Repair of strand breaks by homologous recombination. *Cold Spring Harb Perspect Biol* 2013;5:a012740.
- [64] Schmid-Burgk JL, Höning K, Ebert TS, Hornung V. CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. *Nat Commun* 2016;7:12338.
- [65] Suzuki K, Tsunekawa Y, Hernandez-Benitez R, Wu J, Zhu J, Kim EJ, et al. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* 2016;540:144–9.
- [66] Peters JE, Makarova KS, Shmakov S, Koonin EV. Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proc Natl Acad Sci* 2017;114:E7358–66.
- [67] Strecker J, Ladha A, Gardner Z, Schmid-Burgk JL, Makarova KS, Koonin EV, et al. RNA-guided DNA insertion with CRISPR-associated transposases. *Science* 2019;365:48–53.